AZ Comments on
Docket 2005D-0112

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

General Comments

- In general, the tenor of the document is biased towards OS. While other progression or recurrence based endpoints are discussed, their usage is discouraged through an imbalanced portrayal of the pros and cons versus survival. May of the concerns raised in relation to the use of non-survival endpoints are either equally applicable to survival or can be easily over come using analyses that do not require estimation of the time of recurrence or progression.

- The document proposes several subsidiary analyses of progression-based endpoints in light of concerns relating to uncertainty in the exact timing of progression. Many of these analyses make untenable and potentially biased assumptions regarding how to censor progression times. It is suggested that the document offer one primary analysis of progression and one subsidiary event count analysis that avoids the issue of estimating the time of progression or recurrence entirely.

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
II.A	Lines 79-85 (79-86)	There has been some public discussion concerning whether or not the definition of "adequate and well-controlled" trials meant two randomized Phase III trials. At the podium, FDA (e.g. Dr. Temple) has stated the opinion that one large trial that was unequivocally positive, along with other supportive trial data, could be enough for approval. It has also been clearly pointed out that a Phase III trial with a p-value of 0.045 was not very likely to be enough, unless the supporting data were very strong. Furthermore, FDAMA (1997) states in section 115, CLINICAL INVESTIGATIONS (a) Clarification of the Number of Required Clinical Investigations for Approval, Section 505(d) (21 U.S.C. 355(d)) is amended by adding at the end the following: "If the Secretary determines, based on relevant science, that data from one adequate and well-controlled clinical investigation and confirmatory evidence (obtained prior to or after such investigation) are sufficient to establish effectiveness, the Secretary may consider such data and evidence to constitute substantial evidence for purposes of the preceding sentence." Therefore, text relating to evidence to support drug approval needs to be re-worded to accurately reflect current regulations. Suggest changing to read: "In 1997 the FDA Modernization Act established that data from one well-controlled clinical trial, together with confirmatory evidence obtained either before or after that trial, are sufficient to establish effectiveness. The nature of evidence to support drug approval, including the preferred number of clinical trials, is discussed in general FDA guidance documents. The FDA has

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
		found that evidence from a single trial was sufficient, but generally only in cases in which a single multicenter study provided highly reliable and statistically strong evidence of an important clinical benefit, or in which confirmation of the result in a second trial would be practically or ethically impossible."
II.B	Lines 122-23 (125-127)	The likelihood of a new drug showing clinical benefit via objective response (OR) is not related to the objective response **rate** (ORR), as implied. The ORR merely delineates the likelihood of an individual patient getting an OR if the drug is given to an unselected population. It has nothing to do with whether the OR in an individual patient represents clinical benefit. Response duration is certainly one potential thing to look at in deciding this, also symptom improvement and empirically the degree of shrinkage. Suggest changing to read: "Finally, given the progressive nature of the disease, OR is likely to represent a drug effect which is *reasonably likely* to predict clinical benefit."

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
For accuracy and consistency, the following comments and suggested changes are provided for Line 143, Table 1		
III	Overall Survival	• Regulatory Nature of Evidence column: Suggest changing to read: "Clinical benefit for regular approval" • Assessment column: While the survival date itself is not open to bias, the treatment of the patient between randomization and death is, and since this bias can be applied to both blinded and unblinded trials, suggest changing bullet #2 to read: "Blinding preferred" • Suggest adding a third bullet which reads: "May be biased by any imbalances in treatment decisions" • Some Disadvantages bullet #1: Don't agree that OS requires larger studies per se, but it does require longer follow-up. It only seems to require more patients in order to get a result based on enough survival events early in the trial follow-up period. Suggest deleting bullet 1. • Some Disadvantages bullet #3, suggest changing to read: "Potentially affected by crossover and/or sequential therapy"

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
	Disease-Free Survival	• Some Advantages bullets 1 & 2: As with the statement for OS study size, DFS does not necessarily require fewer patients. Suggest deleting bullet #2 and changing bullet #1 to read: "Clearly a clinical benefit" • Some Disadvantages bullet #1: DFS is validated in breast cancer and colorectal. It can be precisely measured because the appearance of a new lesion is not difficult to determine. Suggest changing bullet #1 to read: "Not a validated survival surrogate in all settings" • Suggest changing bullet #2 to read: "Subject to assessment bias in open-label studies" • Some Disadvantages bullet #3: Not clear on need for this item. Suggest deleting.
	ORR	• Some Advantages: Suggest adding the following as bullet #2: "In a heterogeneous population, an ORR could capture a treatment effect that other time to event endpoints would miss" • Some Disadvantages bullets #1, 2 and 3: Disagree with these bullets. For bullet #2, this statement could equally be surmised in OS and DFS (could be an advantage for ORR). Suggest deleting all and replacing with the following: "May not correlate with changes in other endpoints"
	CR	• Some Disadvantages: Not a significant issue, suggest deleting bullet #2.

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
	PFS	• Regulatory Nature of Evidence: Suggest changing to read: "Clinical benefit for regular approval and surrogate for accelerated approval, depending on the setting" • Assessment, bullet #3: There may be ways to assess PFS in an unbiased way, e.g. assessment of progression status at a fixed time point, such as 6 months or one year. The concept of "clinical benefit" as used in breast cancer is similar (includes CR, PR, and those with stable disease for at least 6 months). This could reduce the documentation and review required. Therefore, suggest changing to read: "Blinded review recommended for open-label studies" • Some Advantages, bullet #3: Suggest deleting. • Some Advantages, bullet #4: Suggest changing to read: "Assessed earlier than survival" • Some Disadvantages: Unclear on the various definitions of PFS exist. Do not agree that it is not a direct measure of benefit (see previous DFS comments). Suggest deleting the first 2 bullets. • Some Disadvantages bullet 3: Suggest changing to read: "Not validated survival surrogate in all settings" (see also previous DFS comment) • Some Disadvantages bullet 4: Suggest changing to read: "Time to Progression (TTP) has to be imputed" • Some Disadvantages bullet 5: Suggest changing to read: "May be subject to assessment bias in open-label studies" • Some Disadvantages: Overstated; suggest deleting 6th and 7th bullets.

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
	Symptom Endpoints	• Regulatory Nature of Evidence: For consistency, suggest changing to read: "Clinical benefit for regular approval • Assessment: Suggest changing to read: "Blinding recommended" • Some Disadvantages 1st bullet: The statement that blinding is often difficult is true for all endpoints. Suggest deleting bullet. • Some Disadvantages 2nd bullet: The idea that missing data are common is not true over short periods. It is only of data are collected long-term that other events happen to the patients which complicate things. Suggest changing to read: "Missing data are problematic" • Some Disadvantages: Suggest changing 3rd bullet to read: "Requires the use of validated instruments" • Some Disadvantages 4th bullet: Overstated; suggest deleting bullet.
III.A	Line 158-9 (162-163)	Disagree with this statement. For longer survival to be a clinical benefit requires that the extra life has some degree of quality. Most of us would not choose to live an extra 2 weeks if we were to be on a ventilator and unconscious all the time or if we would be in pain, semi-comatose and doubly incontinent. Suggest changing to read: "In general, an improvement in survival is a clinical benefit. However, the size of the survival advantage must be weighed against the toxicity of treatment."
	Line 159 (164-5)	As previously commented on OS Assessment in Table 1, bias can be a factor in endpoint measurement where it leads to earlier differences in treatment policies. Anything that could impact the treatment in an arm could introduce bias regardless of endpoint and is covered in previous sentence. Suggest deleting the following; "Bias is not a factor in endpoint measurement."

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
	Lines 177 (183-6)	Suggest adding the following after line 177: It must be noted that there is a potential that increased OS may have arisen from a comparator that has under performed. In this case, improved OS may not always demonstrate a clinical benefit. If the survival increase is small in magnitude it is not necessarily indicative of clinical benefit.
III.B	Lines 189-99 (198-208)	Increasingly, tumor measurements are made on target lesions via imaging which is becoming increasingly sensitive. These measurements are made by radiologists who are or can be unaware of the patient treatment. Bias should not therefore be unavoidable. Suggest including a rationale for the confirmatory evidence needed from a second trial. Also, lines 194-99 (203-8) are repetitive, suggest deleting.
	Lines 201-4 (210-4)	It should not be assumed that the local investigator's assessment would necessarily be biased. This is particularly true if the tumor size measurement is locally performed by a blinded radiologist. We agree with a central review of PFS in open-label trials, but with primary read from the investigator rather than central review. One can use a blinded reviewer at the center. Suggest changing lines 202-4 after "survival or ORR)" to read: "the primary endpoint is assessed by the local center. In a blinded trial, central review is not necessary as there is no opportunity for bias." Suggest that primary tumor response criteria be abandoned in favor of displaying the tumor measurements.
II.B.1	Lines 239-41 (249-52)	Unscheduled visits cannot be avoided in practice. If concerned, one can do an event count analysis and avoid timing altogether. Suggest changing line 241 after "unscheduled assessments" to read: "can be problematic, and may introduce bias."
	Lines 243-45 (253-7)	This approach may introduce a positive bias (i.e., extend the true DFS time.) Suggest changing to read: "The potential effects of bias due to unscheduled assessments can be evaluated by an analysis of the total number of events over the follow-up period regardless of when the events occurred."

Docket No. 2005D-0112. Draft Guidance for Industry on Clinical Endpoints for the Approval of Cancer Drugs and Biologics

Section	Original Line Number (revised version line number)	Comment or proposed replacement text
III.B.2	Line 264-285 (278-94)	In addition to a percentage of patients achieving a certain amount of tumor shrinkage, it is important to display the actual tumor shrinkage achieved for each patient over time. This display is likely to be more informative than a simple binary characterization of response/non-response.
III.B.3	Line 299-300 (311-3)	Don't agree that the definition of progression varies widely. In most patients it will be the appearance of a new lesion. Anyway, in a randomized setting, the definitions of progression are not a significant issue, since they are applied to both groups. Suggest changing to read: "The precise definition of tumor progression is important and should be carefully detailed in the protocol."
III.B.3.c	Lines 345-46 (358-9)	This is not necessary if doing an event count. It is preferred that visits for radiological assessment are symmetrical, however if they are not, alternative statistical methods should be used to address this issue (see line 396)
III.B.3.d	Lines 369-70 (382-4)	The sensitivity analysis suggested should be removed – to censor patient with a known progression event is recommended. Suggest changing to read: "A sensitivity analysis that ignores the imputed timing of event can then be conducted" (see line 396).
	Lines 452-4 (489-91)	No, it's not an effectiveness measure, but it is still a clinical benefit and therefore an endpoint point for registration.

Additional Changes:

Section III.B.3.e. Future methods for assessing progression, lines 398-415 (412-52):

The following section is offered largely in replacement of entire section. Changes to this section are necessary to ensure the 'single time point' approach is properly described. It is helpful, for example, to clarify that this analysis is not at a pre-specified time pre se, but is an 'event count' over the entire follow-up period, up to and including some timepoint following entry of the last patient. Hence, there is no more risk of missing a treatment effect with this approach than there is with the regular log rank analysis of PFS time. Further, work has been submitted for peer reviewed publication which shows an event count analysis produces a result very similar to the log rank analysis of PFS time, which is what would be expected theoretically given the close mathematical link between the two approaches. With respect to linkage to survival, there is no concern per se as survival can and should be collected and evaluated in the normal way. Suggest changing to read:

"In the future, it is important that other methods of progression assessment be evaluated as potential surrogate endpoints for regular approval or accelerated approval. One proposed method (not used to date) is the 'single time point' or, more accurately, an 'event count' analysis which could decrease the complexity of progression assessment and eliminate time-dependent assessment bias. In this analysis, progression would be assessed at a minimum of baseline and at one pre-specified time in follow-up after the last patient had been randomized; typically this would be at the end of the minimum follow-up time specified in the sample size calculation to achieve a desired number of PFS events. The protocol would stipulate that, if a patient progressed prior to the specified time, radiologic scans would be required to document progression. Patients passing through the study without evidence of progression would be required to have a detailed radiologic evaluation at the pre-specified follow-up time. The statistical analysis would compare the number of patients on each study arm with progression on or before the pre-specified time after randomization. In this way the problems associated with the imputation of progression times are avoided entirely. While there is some loss of statistical power, this loss has been shown to be minimal if the proportion of patients with a progression event by the pre-specified follow-up time is not much higher than 75-80%.

Although this approach could provide some advantages and decrease assessment bias, study dropouts prior to progression could present the same difficulty as they do for all progression endpoints. Further theoretical evaluation of this approach is needed. From a more practical standpoint, application of this approach to previously reported trials with PFS as an endpoint would help establish its usefulness and highlight the potential for discrepancy between the approach and the regular analysis of PFS time."

Section IV.B Studies Designed to Demonstrate Non-inferiority, Lines 563-602 (600-72)

Some amends are required to this section. Non-inferiority trials are demanding to design and execute, but remain a valid means of assessing efficacy and safety of a new drug. Also, the true goal of a NI is not as stated in this section and the need for two trials when the aim is to show NI is questioned.

"A randomized trial comparing a new drug to placebo is the most direct and effective way of establishing efficacy and safety of the new drug. However, in oncology, placebo controlled trials are often impossible due to the availability of either approved agents or the use of unapproved, but nevertheless commonly accepted agents. In such circumstances, active controlled, non-inferiority (NI) trials are necessary. The goal of such trials is to demonstrate, indirectly, the absolute effectiveness of a new drug by showing that it would most likely have beaten placebo if placebo controlled trial could have been conducted. A secondary objective of these trials is to examine how well the new drug compares in terms of efficacy and safety, to the active control (ref Wang, Fisher, Carroll). This latter objective is commonly achieved by defining in advance a difference between the new drug and the active control that is to be ruled out statistically. This difference is referred to as the NI margin, and is determined from historical studies of the active control that documented its effect. If the new drug is inferior by more than the non-inferiority margin, then non-inferiority to the degree captured by the margin cannot be established. Previously in oncology drug applications (e.g., Xeloda vs. 5-FU, Cisplatin + taxotere vs. cisplatin + vinorelbine), the NI margin has been arbitrarily set

to be 50% of the active control effect, so that, in these examples, NI was defined as showing at least 50 percent of the active control effect is preserved.

There are several challenges in the design of active-control, NI trials. NI trials necessarily rely on historical data to establish the expected size of treatment effect of the active control. In some situations, the effect of the control drug may not have been established with narrow confidence limits. However, methods do exist that compensate for the level of precision in the active control effect albeit at the expense of the size of the new active control trial, which may need to be extremely large (ref Rothmann). Also, a critical assumption is that the treatment effect of the active control that was observed historically will also be observed in the current population in the new study. This assumption is often difficult to demonstrate unequivocally. Informal comparison of response and death rates on the control arm, of the new active control NI trial with the response and death rates based on historical data may provide some reassurance that this assumption has, or has not, been met. However, it is important to recognize that the performance of the active control is just as much as issue for superiority trials as NI trials; superiority of a new drug to an active control that has grossly under performed can pose difficulties in interpreting whether the new drug has had an true effect, or at least a clinically relevant one. A further problem in NI trials is crossover from the new drug to the control drug, which can bias overall survival toward a showing of no difference. Given the complex issues involved, we strongly recommend that sponsors designing non-inferiority trials consult early with the FDA."

APPENDIX 2: ISSUES TO CONSIDER IN PFS ANALYSIS

Completely missing tumor data, Lines 729-42 (799-828).

"Assessment visits where no data are collected are sometimes followed by death or by assessment visits showing progression; in other cases the subsequent assessment shows no progression. In the latter case, at first glance, it might seem acceptable to continue the patient on study and continue monitoring for evidence of progression. This approach, however, treats missing data differently depending upon subsequent events and could represent informative censoring. Therefore, another possibility is for the primary analysis to include data from subsequent PFS assessments when only a single follow-up visit is missed but censor data when there are two or more missed visits."

This approach is highly problematic and needs revision. It suggests that if a patient has more then one missed visit, they should not be followed further in the trial for disease progression. This would introduce a serious bias in favor of an ineffective drug if visits were missed due to lack of efficacy or undue toxicity. The only sensible approach is to follow all patients for disease progression irrespective of missing visits or, for the same reasons, irrespective of the introduction of additional cancer therapies. The text should therefore be amended to read:

"While missed visits for progression assessment are problematic, all efforts should be made to keep following patients for disease progression irrespective of the number of visits missed. In order to avoid over estimating the true progression time, consideration should be given in the

protocol to simple algorithms for handling a series of missing visits. For example, patients dying without progression, say, 3 months after their last assessment for progression status, might be censored at the time of their last assessment plus 3 months, whereas patients dying without progression within 3 months after their last assessment for progression status would be included with their date of death as the time of progression."

Further, in this same section, it is stated "Reasons for dropouts should be incorporated into procedures for determining censoring and progression status. For instance, for the primary analysis, patients going off-study for undocumented clinical progression, change of cancer treatment, or decreasing performance status could be censored at the last adequate tumor assessment. The secondary sensitivity analysis would include these dropouts as progression events."

This, again, is highly problematic. In the case of true dropouts where patients are lost to follow-up entirely, then the comments above regarding missing visits apply. However, in the case where patients withdraw due to undocumented clinical progression, decreasing performance status or, in particular and critical importance, due to a change of cancer treatment, time to progression cannot be censored in the primary analysis as the censoring mechanism is self evidently informative and so could lead to extremely biased result and, hence, incorrect licensing decisions. For example, suppose drug A is compared to drug B in a trial of 200 patients, 100 per arm. Suppose at the end of the follow-up period there were 30 progression events on drug B (events occurring while the patient was taking the drug) and 50 progression events on drug A. Patients who stop therapy due to undocumented clinical progression, decreasing performance status or a change of cancer treatment are censored in the primary analysis as suggested in the guidance. Based on these results, B is better than A on the primary analysis. However, now suppose that on arm B, 50 patients receive another cancer treatment prior to documented progression due to their deteriorating condition and/or unacceptable toxicity of drug B. Suppose that the corresponding figure for arm A is 10. Looking at all the data, it is clear that A is the better drug.

The guidance therefore offers a fundamentally flawed approach in censoring patients who stop therapy due to undocumented clinical progression, decreasing performance status or a change of cancer treatment. An ineffective, toxic drug where many patients many stop taking therapy due to lack of efficacy and or toxicity, very often early in follow-up, will erroneously appear inferior to a more effective, better tolerated drug where patients remain on therapy longer and long enough to achieve a progression event on therapy.

It is therefore imperative that the guidance be changed to follow an intent-to-treat philosophy as per survival, where follow-up for disease progression continues irrespective of any alteration to randomized therapy and all progression events are counted against randomized treatment irrespective of dropout due to toxicity or other therapy. To continue with the approach outlined would be akin to saying that, for survival, only deaths that occur while patients remain on their randomized therapy should be compared between treatments with all other deaths ignored. Such an approach is clearly inappropriate for survival and is equally inappropriate for PFS.

It is therefore suggested that part starting "Reasons for dropouts should be incorporated into procedures for determining censoring and progression status. The secondary sensitivity analysis would include these dropouts as progression events" is reworded as follows:

"Patients lost to follow-up should be handled in the same way as patients with missing visits. Patients without progression who stop randomized therapy for any reason, for example due to undocumented clinical progression, change of cancer treatment, decreasing performance status or unacceptable toxicity should continue to be followed in so far as is possible for disease progression. Due to the informative nature of events that lead to the cessation of randomized therapy, analyses that censor patients who stop treatment without progression at the last adequate tumor assessment can be biased and misleading and hence can only be considered exploratory in nature."

APPENDIX 3: EXAMPLE TABLES FOR PFS ANALYSIS

In line with the above comments, this Table A, Line 773 (859), which defines the primary analysis of progression, requires amendment. The 3 rows shown should be deleted:

Treatment discontinuation for undocumented progression	Date of last scan of measured lesions	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiologic assessment of measured lesions	Censored
New anticancer treatment started	Date of last radiologic assessment of measured lesions	Censored

The following should then be stated underneath the Table, Line 774 (862-77):

"In line with the intent-to-treat principle underpinning a valid and meaningful analysis of survival, all patients should be followed for disease progression irrespective of any interruption to their randomized therapy. Hence, patients who stop randomized therapy for any reason without progression (i.e., due to undocumented clinical progression, change of cancer treatment, decreasing performance status or unacceptable toxicity) should continue to be followed in so far as is possible for disease progression. Patients who experience a progression event would be included as such in the analysis and those who continue without progression would be censored at their last adequate visit for progression assessment.

The primary analysis, as defined in Table A and incorporating an intent-to-follow patients for progression irrespective of any interruption to their randomized therapy, will therefore compare treatment policies in exactly the same fashion as is standard and common place for overall survival.

Due to the informative nature of events that lead to the cessation of randomized therapy, it is important to recognize that analyses that censor patients who stop treatment without progression at the last adequate tumor assessment can be biased and misleading and hence can only be considered exploratory in nature."